基于 GAN 的手绘草图图像翻译研究综述 *

王建欣 [1a]，史英杰 [1b]，刘　昊 [1c]，黄海峤 [1c]，杜　方 [2]

(1. 北京服装学院 a. 文理学院; b. 商学院; c. 服饰艺术与工程学院, 北京 100029; 2. 宁夏大学 信息工程学院, 银川 750021)

摘　要：手绘草图图像翻译是计算机视觉领域充满挑战的课题，在艺术设计和电子商务领域具有重要的应用价值。目前基于 GAN 的手绘草图图像翻译工作处于起步阶段，文章分析了草图图像翻译面临的挑战性问题，从无控制和精细控制的草图图像翻译两个方面对基于 GAN 的草图图像翻译研究工作进行分析，并对生成图像的评估方法进行总结。基于对已有研究工作的总结归纳，对该领域未来可能的发展趋势进行了展望，为该领域研究人员拓展研究思路提供了线索。

关键词：手绘草图；图像翻译；生成式对抗网络；图像合成；解耦

中图分类号：TP391　　doi: 10.19734/j.issn.1001-3695.2022.01.0027

Research on freehand sketch to image translation based on generative adversarial networks

Wang Jianxin[1a], Shi Yingjie[1b], Liu Hao[1c], Huang Haiqiao[1c], Du Fang[2]

(1. *a. School of Arts & Sciences*, *b. School of Business*, *c. School of Fashion*, *Beijing Institute of Fashion Technology*, *Beijing* 100029, *China*; 2. *School of Information Engineering*, *Ningxia University*, *Yinchuan* 750021, *China*)

Abstract: Freehand sketch to image translation is a challenging subject in the field of computer vision, and has important application value in the fields of art design and e-commerce. At present, sketch to image translation based on GAN is in its infancy. This paper analyzed the challenging problems on sketch to image translation, and summarized the work based on GAN from two aspects of uncontrolled sketch to image translation and finely controlled sketch to image translation. This paper also summarized the method of evaluating generative image. Based on the summary of the existing research, this paper proposed the possible future development trends in this field, which provides clues for researchers to expand their research ideas.

Key words: freehand sketch; image to image translation; generative adversarial network; image synthesis; disentanglement

0　引言

绘画是人类早期的重要艺术活动之一，原始人类可以通过稀疏草图描绘狩猎活动的主要猎物。手绘草图反映了人类大脑对现实世界的视觉感知，任何人都可以通过手绘草图来表达自己的想法并进行辅助交流。从古至今，手绘草图一直是人类可视化物体或场景最直接快速的手段。因此，针对手绘草图的研究在计算机视觉领域很受关注。早期人们对草图的研究主要集中在草图识别、基于草图的图像检索、基于草图的 3D 形状检索等领域，随着深度学习技术的发展，出现了一些新的研究课题，如合成草图、深度草图哈希、实例级草图的图像检索等。近年来图像翻译领域出现了风格迁移和超分辨率等研究成果，手绘草图的图像翻译引起了学术界和工业界的广泛关注。图像翻译是指将一种类型的图像转换成另一种类型的图像，本质上是两个不同图像域之间的相互转换，例如冬天场景图像转换成夏天场景图像，语义图像转换成真实图像，草图转成真实彩色图像等。手绘草图图像翻译是指将人类手绘风格的笔画稀疏、抽象并带有一定噪声的草图转换成既忠实于草图所绘内容又在视觉上具有真实感的图像[1]。传统的基于草图的图像翻译通过图像检索实现[2, 3]：根据草图给定的对象和背景，从大规模图像数据集中搜索与之对应的图像块，再将这些图像块融合在一起，这种方法的缺点是不能生成全新的图像。近年来生成式深度学习尤其是生成式对抗网络(GAN)[4]的迅速发展，使得基于 GAN[4]草图的图像翻译成为可能。由于手绘草图区别于普通图像的特质，导致目前基于 GAN 的手绘草图图像翻译面临着挑战性问题：首先，手绘草图笔画稀疏、抽象，导致手绘草图图像翻译需要矫正变形的笔画和增加更多的细节；其次，一一对应的草图图像数据较少，从而导致训练模型缺乏足够的数据；第三，手绘草图风格多样且难以模仿，导致使用扩充的草图训练的模型不能在真实的手绘草图上泛化。

基于草图的图像翻译可以在实际应用场景中帮助用户创建或设计新颖的图像，是展示人们创造力和交流想法的有效途径之一。在设计领域，草图图像翻译可以帮助设计师快速直观的可视化设计产品。设计师可以通过彩色的线条或者轮廓内填充不完全的彩色块为草图区域指定颜色纹理，草图翻译系统根据这些指导信息生成与其风格相近真实图像，为设计师提供有力的设计参考。在电商领域，草图翻译系统将用户绘制的需求产品草图翻译成真实的商品图像，一方面可帮助用户有效搜索出相似的线上商品，从而增强消费体验；另一方面可为商家分析用户需求提供重要的数据支撑，从而有效促进线上商品的成交量。此外，手绘草图的图像翻译在其

收稿日期：2022-01-13；**修回日期：**2022-03-22　**基金项目：**国家自然科学基金资助项目(61502279，62062058)；北京服装学院重点科研项目(2021A-02)；北京服装学院青年拔尖人才培养计划(YS22-1005096)；北京市服装产业数字化工程技术研究中心科研项目(KJCX20801-30299/016)

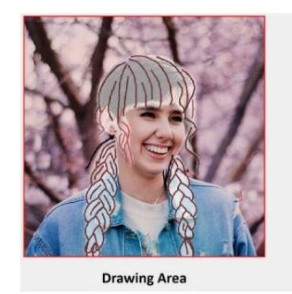
作者简介：王建欣(1989-)，女，河北唐山人，硕士研究生，主要研究方向为时尚大数据分析，图像生成；史英杰(1983-)，女，山东滨州人，副教授，硕导，博士，主要研究方向为云数据管理、时尚大数据管理与分析(shiyingjie1983@163.com)；刘昊(1979-)，男，北京人，副教授，硕导，硕士，主要研究方向为智能交互与可穿戴产品开发；黄海峤(1978-)，男，北京人，副教授，硕导，博士，主要研究方向为服装数字化、服装大数据分析；杜方(1974-)，女，宁夏银川人，教授，硕导，博士，主要研究方向为智能信息检索、大数据管理.

他领域也可大显身手：从稀疏的草图生成逼真的人类面部图像,如图 1[5]所示，可以帮助没有任何绘画基础的目击证人更好的描绘犯罪分子的特征，从而帮助公安机关抓捕犯罪嫌疑人；在影视拍摄领域，编剧或者导演可以根据自己的想象绘制人物角色草图，通过生成逼真的人脸图像对比选择更适合的演员；在图像编辑领域，可以通过草图来编辑人的面部轮廓、头发、胡须、褶皱等，结合风格迁移技术改变妆容肤色，如图 2[6]所示。




图 1　DeepFaceDrawing 的草图翻译效果

Fig. 1　Sketch to image translation effect of deepfacedrawing



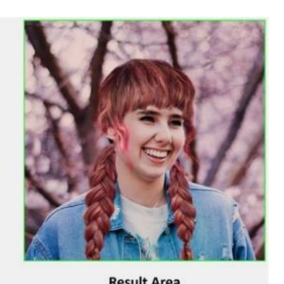


图 2　SketchHairSalon 的草图翻译效果

Fig. 2　Sketch to image translation effect of sketchhairsalon

1　手绘草图图像翻译的挑战

从草图生成逼真的图像并不是一项简单的任务，合成图像需要忠实于给定的草图，同时保持真实性和语义连贯性。手绘草图描绘了对象的近似边界和内部轮廓，是一个特殊的数据域，而真实图像则精确的对应对象的边界并且像素密集，因此手绘草图到图像的翻译是典型的跨模态转换问题。基于 GAN[4]的图像翻译是以数据驱动的，训练过程需要大规模的草图和图像数据，而收集人类手绘草图难度大、成本高，导致可直接使用的草图数据较少，这是基于 GAN[4]的手绘草图图像翻译必须解决的问题。

1.1　手绘草图抽象且多样化

手绘草图是一种生动的数据形式，简洁抽象，而自然图像像素密集，二者有着本质的区别。首先，手绘草图是抽象的，笔画稀疏，色彩单一，非专业绘画人士一般会用比较少的笔画描绘事物；其次，草图是多样化的，不同的人有不同的绘画风格，如图 3 所示，针对同一双鞋子不同人绘制的草图完全不同；最后，手绘草图通常带有一些冗余和嘈杂的笔触，从而使得草图带有一定的噪声。

手绘草图与图像属于不同的数据域，手绘草图图像翻译是跨域模态转换问题，而一般的图像到图像翻译是单模态任务，并且在翻译过程中结合了像素对应[7]类似的硬条件，这使得输出与输入边缘严格对齐。与一般图像翻译相比，手绘草图图像翻译有其自身的特点。首先，手绘草图笔画未与对象边界精确对齐且颜色单调，因此转换过程中需要矫正笔画形变和上色。其次，草图不包含关于背景和细节的更多信息，因此生成模型必须自己插入更多信息。 最后，草图笔画包含的细节特征，模型必须学会处理它们，例如图 3[8]中草图笔画描绘的鞋子表面上的金属装饰。



图 3　手绘草图与自然照片对比

Fig. 3　Hand-drawn sketches versus nature photos

1.2　成对手绘草图数据缺乏

手绘草图图像翻译属于跨模态转换，训练模型需要手绘草图和图像两类数据。表 1 总结了现有草图图像翻译研究工作所使用的数据集，其中包含两种模态的数据集有 Sketchy database[9]、ShoeV2[8]和 ChairV2[8]，其他数据集只包含真实图像或者草图。对于不包含草图的数据，研究人员采用特定方法进行扩充，草图扩充方法如表 2 所示；只包含草图的数据采用草图图像嵌入方法选择与收集的图像最相近的草图作为数据扩充。

表 1　现有草图图像翻译工作使用的数据集

Tab. 1　Datasets used by existing sketch to image translation works

数据集	主题	使用数据集文献
Sketchy database[9]	125 个类别，共包含 12500 个对象的 75471 个草图	文献[1]
CelebA[10]	约 20 万人脸图像	文献[11]
Caltech-UCSD Birds-200-2011[12]	11.7 千张鸟类图像	文献[11]
斯坦福汽车数据集[13]	16 千张汽车图像	文献[11]
Flickr-Faces-HQ (FFHQ)[14]	7 万张肖像图像	文献[15]、[16]
CUFS[17]	606 张人脸和对应的素描草图	文献[18]
Celeb A Mask-HQ[19]	3 万张高分辨率人脸图像及人脸属性分割蒙版	文献[5]、[20]
CelebA-HQ 数据集[21]	3 万张肖像图像	文献[22]、[23]
COCO Stuff[24]	91 个 stuff 类，164 千个图像及注释	文献[25]
Tuberlin 数据集[26]	250 个类别，2 万张草图	文献[25]
QuickDraw[27]	345 个类别，5000 万张草图	文献[25]
ShoeV2[8]	6648 张草图 2000 张图像	文献[28]
ChairV2[8]	1297 幅草图 400 张图像	文献[28]
SketchyCOCO[25]	17 类 6 万对以上的草图和图像	文献[29]
Oxford-102 数据集[30]	102 个花类，每类 40 至 258 张图像	文献[31]

1.3　人类手绘草图模仿困难

目前公开数据集中成对的草图图像数据较少，一些研究工作雇佣人工绘制草图[5, 32]，通常此类方法的草图图像翻译效果较好，然而人工绘制草图的成本比较高，并不适用于大规模的草图数据集生成。为此，研究人员提出各种方法来扩充草图数据，然后使用扩充的草图和图像进行训练，如表 2 所示。扩充草图的方法可分为三类：提取真实图像的边缘图作为草图，如使用整体嵌套边缘检测(HED)[33]、XDoG[34]边缘检测器、FDoG[35]过滤器等，此类方法获得的草图细节依赖于阈值大小；使用图像草图翻译网络生成草图，如 Im2pencil[36]、Photosketching[37]，此类方法生成的草图能够很好的捕捉目标轮廓，甚至精细描绘，但不能模仿普通用户的稀疏抽象的手绘草图；抽象笔画来模仿手绘草图，如对边缘图的笔画进行随机变形或者简化线条以去除重复、潦草的边，这类方法对原有的线条笔画做比较小的改动。总的来说，目前已有的草图扩充方法或者直接提取边缘图作为草图，或者利用草图翻译网络生成草图，然而这些草图不能模拟新手用户稀疏的笔画，研究新的草图扩充方法或者提升模型到手绘草图的泛化能力是草图图像翻译的重点问题之一。

表 2　草图扩充方法

Tab. 2　Methods of sketchy database augmentation

类别	草图扩充方法	特点	代表工作
提取真实图像边缘作为草图	HED[33]	优点：能够比较完整的提取对象的轮廓 缺点：与对象边界精确对齐，包含太多背景信息	文献[1]
	二值化、腐蚀等	优点：可以减少边缘过多的细节笔画 缺点：不能形成变形的笔画来模仿稀疏的草图	文献[1]、[38]
	XDoG[34]边缘检测器	优点：边界清晰 缺点：包含一定的细节笔画，与对象边界对齐	文献[11]、[18]
	Photoshop 影印[39]	优点：能够提取比较完整的边界信息 缺点：笔画不够连续，包含比较多的阴影细节信息	文献[5]、[11]、[16]、[18]
	FDoG[35]过滤器	优点：轮廓信息明显 缺点：包含一定的细节阴影和笔画，与边界对齐	文献[11]
	语义图边界	优点：不与对象边界精确对齐 缺点：只包含语义图的边界线条	文献[20]
	Sketch master[40]	优点：能够比较完整的提取对象的信息 缺点：包含过多的细节阴影，与对象边界对齐	文献[41]
	基于离散 H 通道的边缘检测	优点：可以提取服饰图案和饰品的边缘 缺点：与对象边界精确对齐	文献[38]
网络生成草图	Im2pencil[36]	优点：能够比较完整的提取对象的信息 缺点：不能模仿笔画稀疏且变形的草图	文献[42]
	Photosketching[37]	优点：笔画稀疏且有一定的变形 缺点：部分笔画断断续续，不够连贯	文献[18]
	无监督草图生成网络 TOM	优点：可以生成 10 种不同风格的草图 缺点：轮廓笔画变形不是很大，部分线条不连贯	文献[23]
抽象笔画	笔画变形[43]	优点：可以模仿人类草图不与对象边界对齐 缺点：与人类手绘草图的笔画风格有一定差距	文献[20]
	笔画简化[44]	优点：使边界线条更加清晰，去除过多的阴影细节 缺点：部分线条不连贯	文献[11]、[16]、[18]、[22]、[42]
	移动最小二乘对轮廓变形	优点：能够实现笔画一定变形的手绘草图风格 缺点：不能模仿笔画过于夸张变形的草图	文献[31]
	积分矢量场[45]模拟笔画	优点：可以生成手绘风格的笔触 缺点：笔画比较密集，风格单一	文献[46]

2　基于 GAN 的草图图像翻译方法

草图图像翻译的目标是学习草图到图像的跨域图像映射，根据对生成图像的控制程度，可将已有的研究工作分成两类：一类是无控制的草图图像翻译，目前大部分方法是利用配对数据或未配对数据的条件生成对抗网络(CGAN)[47]解决问题。另一类是精细控制的草图图像翻译，从草图到图像的映射本质上是多模态的，为了实现对输出进行精细控制，研究人员提出了使用属性和笔画控制输出的图像。

2.1　无控制的草图图像翻译

草图到图像翻译旨在学习两个不同图像域之间的转换。按照训练方式的不同，一般的草图图像翻译可以分为基于监督的方法和无监督的方法两类，如表 3 所示。通用的图像翻译框架要求成对的草图和图像，使用条件 GAN 对配对图像进行一对一映射，此为监督学习的方法。无监督的基于 GAN 的草图图像翻译方法使用一对 GAN 将图像从源域映射到目标域，然后再将其返回到源域，允许使用未配对的数据进行训练。

2.1.1 *基于草图监督的方法*

Pix2pix[7]是一个通用的图像翻译框架，常被用来作为基线对比。但它不是专门针对草图设计的，只有专业的写实草图甚至边缘图作为输入时才能产生合理的结果，其翻译过程是推断笔画之间缺失的纹理或阴影信息，因此当使用稀疏的手绘草图作条件时，网络不能产生很好的结果。图像到图像生成模型通常无法用于草图图像生成，因为草图和图像之间的域差距很大，无法直接在视觉空间中进行逐像素对齐。Pix2pixHD[48]也是图像到图像的转换方法，可以生成分辨率为 2048×1024 的图像，但它同样不能处理手绘草图问题。

手绘草图作为一种通用的表达方式，其所描绘的内容包罗万象。根据草图翻译生成的图像对象可以分为生成多类别的图像、生成发型人脸和生成场景级图像，下面分别对这三类方法进行具体介绍。

1)生成多类别的图像

2018 年，James Hays 等人提出了 SketchyGAN[1]，它训练以草图图像对的类标签为条件的编码器-解码器模型，是一种基于 GAN 的端到端的多模态合成方法，可以生成马、沙发、摩托车等 50 个类别的对象。在生成器和判别器使用屏蔽剩余单元(MRU)块来代替卷积层，通过掩码输入不同比例的图像金字塔提取特征。同时为了鼓励生成图像的多样性，作者提出一种多样性损失，最大化具有不同噪声向量的两个相同输入草图的输出之间的 L1 距离。同年，Yongyi Lu 等人提出了另一种解决方案，ContextualGAN[11]。他把草图到图像转换问题，转换成草图作为上下文弱约束的图像补全问题。通过使用联合图像来学习草图和相应图像的联合分布，避免跨域学习中的复杂问题，这种方法也可用于图像到草图的生成。文献[29]提出了从草图到边缘图再到图像的两阶段草图图像翻译模型，通过引入特征间相关性学习可以使模型在无类别标签下生成与类别一致的图像。为了帮助新手用户创建草图

对象，Arnab Ghosh 等人提出了 iSketchNFill[42]。它是基于交互式 GAN 的草图到图像的翻译系统，引入了一种基于门控的类调节方法从单个生成器网络生成篮球、鸡肉、饼干、纸杯蛋糕等 10 类图像。当用户绘制所需对象类型的草图时，系统会自动推荐笔画反馈给用户帮助其完成草图，并根据类条件进行纹理填充。它由基于非图像生成网络的形状完成阶段[49]和基于 MUNIT[50]的编码器-解码器模型的类条件外观转换阶段组成，可以生成 256×256 分辨率的图像。

总的来说，生成多类别的图像往往需要大量的训练数据，以上三种方法都提出了不同的草图数据扩充办法。但是他们扩充的草图更加接近于真实图像的边缘图，当使用稀疏抽象的真实人类手绘草图时往往不能生成合理的图像，如图 4[51]所示。此外，生成图像的分辨率较低，如 SketchyGAN[1]只能生成 64×64 分辨率的图像。

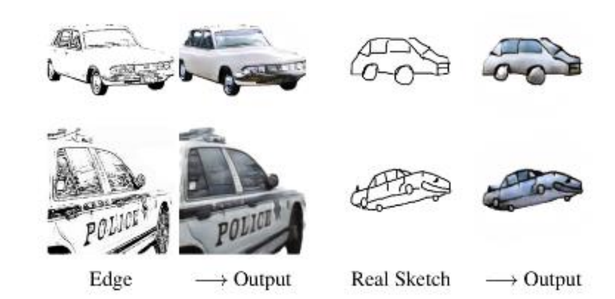


图 4　使用边缘图模拟草图与真实草图到图像翻译效果对比

Fig. 4　Comparison of the effect of sketch to image translation using edge maps and freehand sketches

表 3　无控制的草图图像翻译方法

Tab. 3　Methods of uncontrolled sketch to image translation

训练方式	代表工作	主题	图像分辨率	特点
监督	SketchyGAN[1]	马、沙发等 50 个类	64×64	优点：多尺度输入，提出了新的数据增强技术 缺点：分辨率低，不够真实或忠实于草图
	ContextualGAN[11]	人脸、鸟和汽车	64×128	优点：使用联合图像来学习草图和相应图像的联合分布 缺点：对人脸草图忠诚度较低，分辨率低
	文献[29]	长颈鹿、大象等 14 个类	256×256	优点：两阶段的草图图像生成模型，能够生成多种类别的图像 缺点：生成的图像不够真实，部分类别不能生成符合草图的图像
	iSketchNFill[42]	篮球、饼干等 10 类	256×256	优点：可交互，软门控机制实现类调节 缺点：分辨率较低，数据集大部分为圆形
	HIS[32]	发型	512×512	优点：设计自增强模块生成纹理细致的发型， 缺点：不能生成辫子等复杂的发型，生成的发型不够真实
	Cali-Sketch[18]	人脸	256×256	优点：将跨领域转换问题解耦为两步，笔画校准和图像合成 缺点：训练数据集小，不能泛化到笔画抽象的草图
	文献[41]	风景、人脸、包、鞋	256×256 64×64	优点：增加隐码向量实现多模态输出 缺点：数据集小，扩充的草图接近边缘图，模型参数多
	DeepFaceDrawing[5]	人脸	512×512	优点：利用流形投影来提高手绘草图的生成质量和鲁棒性 缺点：不能修复组件布局中的错误，无颜色控制
	DeepFacePencil[20]	人脸	256×256	优点：设计 SAP 自适应地处理失真的笔画 缺点：分辨率低，对新手画家草图生成效果不够真实
	pSp[22]	人脸	512×512	优点：可应用到多种图像翻译任务中，实现人脸侧面草图图像翻译 缺点：依赖于预训练的生成模型
	SketchyCOCO[25]	场景级	256×256	优点：训练阶段不使用手绘草图，提取属性向量作为桥梁生成图像 缺点：数据集有偏差，抽象的草图不能分割
	文献[31]	猫、花卉	256×256	优点：变形边缘图模拟草图，使用第二层 GAN 网络丰富纹理特征 缺点：无法将夸张的变形笔画转换为真实感的图像
	文献[38]	民族服饰	256×256	优点：提出针对民族服饰特点的边缘检测方法和草图模拟方法 缺点：数据集小，部分生成的图像色彩细节模糊、变化突兀
无监督	US2P[28]	鞋子、沙发	128×128	优点：不需要成对的训练数据，多模态输出 缺点：图像分辨率低，耗费算力

2)生成发型人脸

毛发模拟是计算机图形学的一个非常具有挑战性的研究课题，因为它往往需要对数十万根毛发进行模拟，同时要考虑毛发之间的运动特性和相互碰撞。随着生成式深度学习的发展，研究人员将目光投放在基于 GAN 的毛发生成。HIS[32]提出了基于 GAN 的草图到发型转换的两阶段模型，同时构造了 640 对头发面积限制在 512×512 以内的高分辨率的发型草图图像数据。该模型能够输入发型草图或者低分辨率的头发图像产生逼真头发图像。具体来说，首先将发型草图或者低分辨率的头发图像应用 Pix2pix[7]框架生成粗糙的发型图像，然后将粗糙发型图像输入具有自增强能力的再生网络生成高质量的结果。其中的自增强能力是作者提出的结构提取层，从头发图像中提取纹理和方向图，从而生成更精细的纹理和发丝。

人脸相关问题一直都是计算机视觉应用领域研究的重点，如人脸识别、人脸检测等。同样在生成式深度学习的各项任务中，合成人脸一直都是研究人员关注的热点。Weihao Xia 等人提出了 Cali-Sketch[18]，其是基于草图的人像合成的两阶段网络。具体来说，第一阶段笔画校准网络负责将输入的稀疏的草图转换为更详细和校准的类似于边缘图的草图。第二阶段将精制的草图用于图像合成网络，以获得逼真的肖像图像。文献[41]使用隐码向量来实现人脸图像多模态的输出，但是图像分辨率仅为 64×64。为了解决过度拟合草图的问题，Lin Gao 等人提出了 DeepFaceDrawing[5]，能够生成分辨率为 512×512 的逼真图像。其实验采用高清人脸数据集，并通过对人脸图像 PS 影印[39]加笔画简化[44]的方法扩充草图。为了

从粗糙稀疏的或不完整的草图也能够生成高质量的面部图像，作者将扩充的草图作为软约束。具体是采用局部到全局方法，将人脸分为左眼、右眼、鼻子、嘴和面部剩余五个关键人脸组件部分，学习这些组件的特征嵌入。然后训练深度神经网络将嵌入的组件特征映射到逼真的图像，同时使用流形投影来提高手绘草图的生成质量和鲁棒性。Yuhang Li 等人提出了另一种解决方案，DeepFacePencil[20]。它使用一个名为空间注意力池(SAP)的模块，可以自适应地调整生成图像的真实性和生成图像与输入草图之间的一致性之间的空间变化平衡。其网络使用双生成器框架，来促进 SAP 感知局部不够真实完美的笔画，并将合成的面部区域从不完美的笔画修正为逼真的图像域。pSp[22]是一个通用的图像翻译框架，它将编码器与 StyleGAN2[52]解码器相结合，可应用于草图到图像的转换，且能够实现多样化的输出，不止生成正面人脸图像。但是草图几何被编码在潜在代码中，由 pSp[22]生成的人脸通常不会忠实地尊重输入草图，它采用的风格混合操作也会不利地影响合成真实几何形状的面部。

总的来说，目前绝大部分工作是生成正面人脸图像，利用人脸的固定结构可以生成高质量的图像。未来，探索其他属性比如头部姿势和照明，如何克服草图语义的模糊性生成准确的头发、背景、颈部等的边界是具有挑战性的工作。

3)生成场景级图像

与单个对象的图像不同，场景级的图像结构复杂，涉及多个对象和复杂的背景关系。为此，Chengying Gao 等人提出了 SketchyCOCO[25]，专注于从手绘草图生成整个场景的图像。由于草图绘制的粗糙程度不同，它将草图分为前景和背景两部分顺序生成图像。前景是指论文数据集中的鹿、斑马、大象等动物，背景是指草地、蓝天、树木等。前景生成目的是尽可能符合用户的要求，背景部分生成与草图对齐。针对前景草图的抽象性和差异性，作者设计了新的神经网络算法 EdgeGAN，在训练阶段不需要成对的手绘草图和图像而仅使用图像以及对应的边缘图。具体做法是将前景和对应的边缘图输入网络，学习图像和边缘图的公共属性向量表示，最后通过输入草图的属性向量映射到对应的图像。背景部分的图像生成则用 Pix2pix[7]架构完成，把生成的前景图像和背景草图一起送给网络可以生成分辨率为 128×128 和 256×256 的场景级图像。

草图到场景级图像的合成工作比较少，现有的工作生成的图像分辨率较低。对于数据集构建的相关技术问题，依赖于更先进的草图分割技术来处理抽象的草图。

此外，文献[31]使用双层级联的 GAN 网络来生成分辨率更高纹理丰富的图像，可以生成猫类、花卉类图像。针对手绘草图稀缺的问题，作者提出了移动最小二乘的策略来对提取的边缘图轮廓进行变形来模拟手绘草图的风格。文献[38]专注于中国少数民族服饰的草图图像翻译，针对服饰特点设计服饰图案轮廓提取方法，并根据草图风格的特点对边缘图处理以模仿草图。总的来说，以上两种方法生成的图像还不够真实，无法处理带有密集笔画或者夸张线条的草图。

2.1.2 基于草图无监督的研究方法

由于成对的数据获取难度大成本高，研究人员开发了一系列无监督的方法来实现图像翻译。在通用的图像翻译领域，CycleGAN[53]是基于无监督的图像翻译方法，之后 MUNIT[50]将图像数据分为内容部分和风格部分，从不同的数据空间采样进行重构实现图像域之间的多对多映射；U-GAT-IT[54]提出一个注意力模块引导注意力图区分源域和目标域，AdaLIN 函数引导模型更加灵活地控制形状和纹理的变化。以上方法都不是专门针对草图图像翻译的方法，无法有效处理稀疏的几何变形的人类手绘草图。

US2P[28]是采用不成对的草图图像数据的两阶段无监督模型，同时可以生成多样化的逼真的图像。首先通过循环一致性损失[53]的监督将输入草图转换为灰度图像，然后利用单独的 GAN 模型进行基于样本的着色，下面具体介绍这两个阶段。

第一阶段进行形状翻译，用来处理草图的空间形变，包括抽象线条和多变的绘画风格。此阶段使用的数据是未配对的草图和灰度图，包括草图到灰度图和灰度图到草图两对映射，使用循环一致性损失监督，类似于 CycleGAN[53]的模型。针对草图的特殊性，存在密集的无用笔画或者细节噪声而引入了自监督和注意力模块。自监督模块用来将噪声草图恢复成原始的干净草图，如图 5[28]所示。由于草图空白的区域大，使用注意力模块来重新加权注意力图来抑制激活密集笔画区域，进而忽略噪声干扰，如图 6[28]所示。第二阶段称为内容丰富，网络将灰度图生成包含细节的彩色图像。此阶段使用配对的灰度图和图像进行训练可以提供参考图像作为样式指导，并遵循 AdaIN[55]通过调整特征图来使输出多样化。

由于形状转换网络是双向的，从草图转换为灰度图和从灰度图转换为草图，所以 US2P[28]可以将图像转换为草图，还可以应用到基于草图的无监督检索。总体上来说，US2P[28]只关注鞋和沙发两类数据，且草图数据量较少，生成的图像分辨率仅为 128×128。由于成对的手绘草图图像很难获得，未来，突破循环一致性损失的瓶颈，探索更先进的无监督方法是解决草图图像翻译难点。

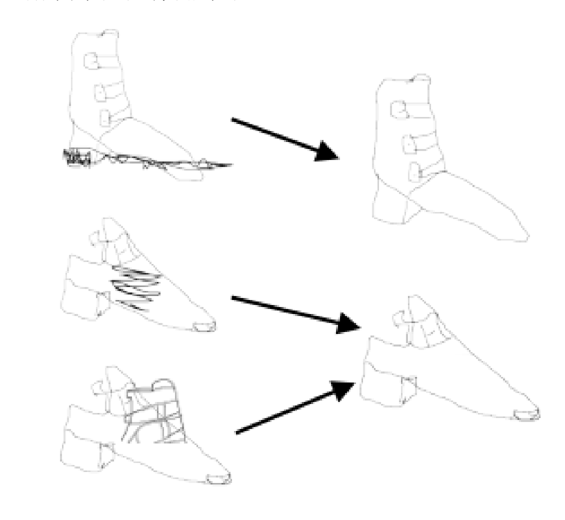


图 5　自监督模块去噪

Fig. 5　Denoising by self-supervision

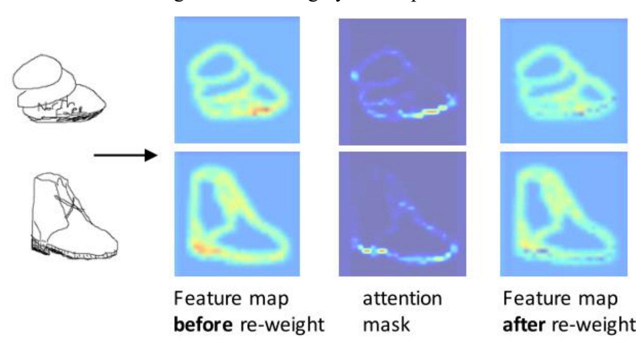


图 6　注意力模块去噪

Fig. 6　Denoising by attention

2.2　精细控制的草图图像翻译

虽然以上部分工作支持多模态的图像生成，但是生成图像的属性、风格等不可控。为了让用户更好的对输出进行精细控制，研究人员进一步提出了一系列方法，如表 4 所示。已有研究工作控制的具体对象包括图像属性和笔画。

表 4　精细控制的草图图像翻译方法

Tab. 4　Methods of finely controled sketch to image translation

代表性工作	主题	图像分辨率	交互性	交互方式	精细控制	风格化	算法特点
BHS[46]	发型胡须	512×512	有	整合矢量场，转换为引导笔划	毛发结构和颜色属性控制	是	笔画引导生成面部毛发，具有交互能力
MichiGAN[15]	发型	512×512	有	参考模式和绘制模式	形状、结构、外观和背景属性解耦	是	使用方向图来实现结构控制，头发外观作为一个风格转移问题
SketchHairSalon[6]	发型	512×512	有	自动完成重复的未编织笔画、给定编织结	彩色草图指定发型的形状、结构和外观	无	彩色笔画引导发型颜色，采用带有自注意力模块的编码器-解码器生成器[56]来忠实于草图
DeepFaceEditing[16]	人脸	512×512	有	草图编辑，给定参考图像改变弯外观	几何和外观解耦	是	局部解缠模块将局部区域的图像和草图嵌入共享空间、全局融合模块
SSS2I[23]	人脸、艺术绘画	1024×1024	无	无	内容和风格解耦	是	自监督的自编码器 AE[57]生成草图，动量[58]互信息最小化更好的解耦，改进 DMI[59]
Sketch Your Own Gan[60]	马、猫和教堂	256×256	无	无	草图控制形状和姿势	无	使用较少的草图创建生成模型，利用预训练的生成模型，设计调整模型权重来匹配草图

2.2.1 图像属性控制

图像属性控制是指将需要翻译的图像分解为几个视觉属性，对每一个属性，设计相应的模块进行控制生成图像。其中，图像的纹理风格能更好的帮助用户指定期望的目标，为此，研究人员做了一些基于范例的图像翻译方法的研究工作。

基于范例的图像翻译是指将图像(如语义分割图、人体骨骼关键点、边缘图等)按照指定风格(颜色、纹理等)参考图像进行图像翻译。网络接收源域图片时也接收一张与其具有相似语义信息的目标域的范例图片，它具有用户期望的目标风格，网络将这两个图像同时作为条件学习输出符合指定风格的图片。CoCosnet[61]提出一个基于范例的图像翻译的框架，方法是建立输入图及范例图之间的密集语义对应，以此定位输入图在范例中相应位置的颜色和纹理信息，使生成的图像风格与范例中物体对应，可以应用到图像编辑和人脸上妆。RBNet[62]利用参考图像给草图或者边缘图上色。此外，文献[59]提出一种基于艺术风格范例的草图图像翻译方法，可以生成分辨率为 512×512 高质量图像。其采用的是 SketchyGAN[1]中的数据扩充方法构建草图，同时论文中展示了网络也可以生成人体图像。但以上方法都不是针对人类手绘草图生成真实图像的方法。

为了实现可控的头发操作，MichiGAN[15]提出了交互式人像头发图像生成方法，专为以解耦属性(包括形状、结构、外观和背景)为条件的肖像照片生成逼真的头发图像。其交互式系统还可以通过参考人像或者绘画对图像进行局部和细节的编辑。文献[46]也是一种交互式方法可以合成图像中头发和胡须。DeepFaceEditing[16]是 Lin Gao 等人的最新成果，是一种专为人脸图像设计的结构化解缠框架，通过几何和外观的解缠控制来实现人脸生成和编辑。具体做法是采用局部到全局的方法来合并人脸，局部组件图像被分解为几何和外观表示，最后在对其进行全局融合，最终生成高质量的图像。它的原理是利用草图提取几何表示，因此支持通过草图编辑人脸图像。由此产生的方法既可以从人脸图像中提取几何和外观表示，也可以直接从人脸草图中提取几何表示。Bingchen Liu 等人提出了 SSS2I[23]，是一种基于范例的带有手绘草图的图像合成方法。为解决成对手绘草图图像的缺失问题，作者提出一种基于 GAN 的域转移无监督模型 TOM。模型将草图合成视为由 RGB 图像域 R 映射到线草图域 S 图像域转移问题，通过在线特征匹配为每个图像合成多个草图。以风格范例为导向的草图到图像生成主要由两部分组成，第一阶段把草图转换为彩色图像，第二阶段使用对抗网络进一步细化彩色图片细节，提高分辨率和合成质量。首先使用合成的配对数据，通过自监督的自编码器(AE)[57]来将草图和 RGB 图像的内容和风格特征分离。具体来说，先把图片进行风格编码提取风格特征,然后把草图进行内容编码提取内容特征，通过一个简单的风格分类器来让提取后的风格和内容进一步解耦，然后将二者输入给生成器，将草图转换为图片。再把转换后的图片经过另一个生成器,进一步的细化图片的分辨率和风格。

2.2.2 笔画控制

在毛发生成方面，研究人员认为带有颜色的笔画能够为图像生成提供属性指导。BHS[46]使用一组类似草图的“引导笔画”来描述要合成的头发的局部形状和颜色，同时更加方便交互。编辑一个提取毛发信息的矢量场，使用相对较少的用户输入调整发型的整体结构，通过合成的引导笔划来简单地编辑、添加或删除单个笔画来实现最终图像形状和颜色的细微局部变化。Hongbo Fu 等人认为彩色头发草图已经隐含了目标头发形状和头发外观信息，为此，作者提出了 SketchHairSalon[6]一个新颖的网络框架。该框架可以直接从一组彩色笔画合成 512×512 分辨率的逼真头发图像，它由使用了自注意力模块的草图到亚光(S2M-Net)生成和草图到图像(S2I-Net)生成两部分网络组成。同时为了训练网络，作者构建了一个新数据集，包含数千个带人工注释的头发草图图像对和相应的头发遮罩。其设计界面如图 2 所示，包括头发结构定制、头发形状优化、头发的外观定制、自动完成草图等功能。由于训练高质量的生成模型需要大规模的数据集和高性能的计算平台，且训练耗时通常较长。文献[60]提出了一种用少量草图示例定制生成模型的方法，利用在大规模数据上预先训练的现成生成模型，通过草图来指定对象的形状和姿势，同时保持真实性和多样性。其原理是设计一种跨域模型微调的方法来调整模型权重的子集以匹配用户草图，使新模型创建类似于用户草图的图像，同时保留预训练模型的颜色、纹理和细节。

目前的工作主要是对毛发和人脸两个任务做精细控制，算法针对性强，不适用于其他任务的草图图像翻译控制。尤其在艺术设计领域，精细的控制生成的图像或者图像编辑能够辅助设计师进行设计，具有非常好的商业价值，同时也非常具有挑战性，是未来很有前景的研究方向。

3　结果评估

评估生成模型的性能是一项复杂的任务，由于一些定量

指标缺乏与人类感知的一致性[63]，许多研究工作仍然依赖于定性的人工评价评估合成图像的质量。对于特定任务或应用程序，评估不仅应基于最终图像质量，还应考虑生成的图像与条件输入的匹配程度，以及服务于预期应用程序或任务。基于 GAN 的手绘草图图像翻译的结果评估主要包含定性评估和定量评估两类，如表 5 所示。

a) 定性评估。常用的定性评估有感知研究、可用性研究、泛化能力比较、消融研究和与先进模型比较等方法。感知研究是邀请一些没有受过专业绘画训练的人员来评价生成的图像，通常以在线问卷的形式让他们对生成的图像进行评估，然后进行投票或者分数统计。可用性研究也是邀请一部分用户实地体验草图翻译系统，然后填写问卷来评估可用性和有效性。泛化能力比较是训练好模型后，使用稀疏的或者夸张变形的没有绘画经验的人绘制的草图测试模型生成结果，通常此类模型训练时采用的数据多为边缘图或者接近边缘图的专业手绘草图。以上的定性评估方法是最直接最有效的评估方式，也最能真实的反映模型生成图片的质量。

表 5　草图图像翻译评价指标

Tab. 5　Evaluation index of sketch to image translation

评估方法	指标	含义	代表工作
定性研究	感知研究	真实度	文献[1]、[5]、[16]
		忠诚度	文献[1]、[5]、[16]
		自然度	文献[6]
	可用性研究	用户测试系统，有用性、有效性	文献[5]、[6]
	泛化能力比较	不同变形程度的合成草图、手绘草图	文献[11]、[16]、[20]
	消融研究	研究模型组件的有效性	文献[1]、[16]
	与先进模型比较	对比模型效果	文献[1]、[5]
定量研究	FID[64]	测量分布相似性	文献[23]、[25]
	风格相关性 SR[61]	衡量颜色和纹理的一致性	文献[23]
	LPIPS[65]	使用神经网络评估感知相似性	文献[23]
	IS[66]	计算分布的 KL 散度	文献[20]
	L2 Gabor feature[67]	评估相似度	文献[25]
	SAD[68]	绝对差异总和	文献[5]
	IoU	评估边界区域准确性	文献[5]
	SSIM[69]	评估相似度	文献[11]、[25]
	AMT	识别真假	文献[42]

b) 定量评估。研究表明，仅选择一种指标来证明模型的有效性通常是不够的，一般模型都使用以下指标的组合来更有效地衡量其性能。Fréchet Inception Distance[64](FID)量两组之间的分布相似性，并作为生成图像的多样性和质量以及图像与草图的匹配程度的度量。较低的 FID 表示生成数据的分布更接近真实样本的分布。结构相似性指数度量 (SSIM)[69]给出图像与参考图像的相对相似性分数，其中较低的分数表示生成图像的多样性较高(即模式崩溃较少)。学习感知图像块相似性(LPIPS)[65]使用从神经网络中学习到的深度特征来评估图像块之间的感知相似性。Inception Score[66](IS)是应用在 ImageNet 数据集上预训练的 Inception 模型来提取生成图像的特征，并计算条件类分布和边缘类分布之间的 KL 散度，更高的 IS 呈现更高质量的生成图像。风格相关性(SR)[61]是利用低级感知特征的距离来衡量颜色和纹理的一致性。它检查模型与输入的风格一致性，并反映模型的内容或者风格分离性能。形状相似度 L2 Gabor feature[67]和结构相似性度量(SSIM)[69]是用于评估生成的图像和真实图像的相似度的一种度量。文献[5]针对草图生成头发使用绝对差异总和(SAD)[68]以评估头发磨砂生成的准确性，同时使用联合交集(IoU)对生成的遮罩和地面实况进行阈值处理来评估边界区域的准确性。

部分评估指标展示了有效性，但是不同的评估方法适合于不同的模型。例如 Inception Score[66](IS)评估图像有局限性，且分数高低不能如实反映图像的真实度。Fréchet Inception Distance[64](FID)可评估与 ImageNet 不同的数据，但它们都不能反映过拟合的问题。SSIM[69]在图像去噪、图像相似度评价方面表现较好，是一个广泛使用的图像质量评价指标。

4　结束语

基于 GAN 的手绘草图图像翻译通过手绘草图指定合成目标，从而实现可控制的图像生成。在实际应用中，可以根据特定要求生成图像。本文首先分析了手绘草图图像翻译面临的挑战，并对相关工作和评价指标进行了总结和分析。目前基于 GAN 的手绘草图图像翻译已有一些研究工作，但仍处于起步阶段。人类手绘草图复杂多变，描绘对象千变万化，仍有很多有价值的问题亟待解决。

a) 人类手绘草图数据扩充。由于缺乏草图和图像的大规模数据集，收集手绘草图又非常耗时。而且针对不同描绘对象的草图图像翻译通常需要不同的数据集，需要大规模的数据集训练模型。现有的数据增强方法如基于全图的增强(旋转，移位)，或者笔画变形、笔画加粗，都没有考虑如何模仿人类的真实绘画风格[70]。文献[23]探索了一种无监督的方法合成草图，解决了草图数据缺乏的问题，但其合成的草图更类似于专业的写实风格。如图 7[51]所示，采用合成草图训练的模型无法在真实草图上泛化，因此如何合成模仿人类多种真实绘画风格的草图，并缩小合成草图和真实草图之间的域差距[51]，是未来研究工作的重点和难点。

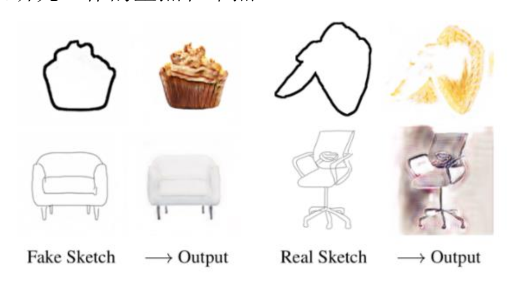


图 7　合成草图与真实草图到图像翻译效果对比

Fig. 7　Comparison of sketch to image translation effect used by synthetic sketches and freehand sketches

b) 精细控制生成的图像。尽管大量的工作支持多模态的草图图像翻译，但是具体纹理、颜色、材质特征等很难控制。基于范例的草图图像翻译可以通过指定单个风格范例图片来控制生成图像的纹理和颜色等信息。未来，参考多风格范例或者使用带有颜色的笔画来控制生成的图像的工作更具有商业价值，比如可以减少动画、电影和视频游戏故事板中的重复工作。在艺术设计领域，如何表现物体的材质属性而不单单是颜色，从而更好的辅助设计师进行创作也是未来很有探索价值的方向。

c) 草图到艺术风格图像生成。目前，大多数研究工作都集中在从草图合成逼真的自然照片图像，艺术图像与其他类型图像的区别在于艺术风格的多样性，这些艺术风格会影响草图如何合成为全彩色的纹理图像。文献[59]研究了基于草图的艺术风格(例如，印象派、现实主义等)图像合成，局限是某些艺术风格的特征很难被模型学习，不能很好的平衡模型

从草图的语义特征和风格参考图像中学习表示。将草图转换为艺术绘画风格的图像为推动深度神经网络在捕捉和翻译各种艺术风格方面的工作作出贡献。未来，此项工作不仅可以用作娱乐应用，能够让用户体会艺术绘画创作的乐趣，提升艺术修养，而且可以从多个艺术风格合成图像，辅助艺术家进行创意艺术创作。

参考文献：

[1] Chen Wengling, Hays J. Sketchygan: towards diverse and realistic sketch to image synthesis [C]// Proc of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 9416–9425.

[2] Chen Tao, Cheng Mingming, Tan Ping, *et al.* Sketch2photo: internet image montage [J]. ACM Transactions on Graphics, 2009, 28 (5): 1–10.

[3] Eitz M, Richter R, Hildebrand K, *et al.* Photosketcher: interactive sketch-based image synthesis [J]. IEEE Computer Graphics and Applications, 2011, 31 (6): 56–66.

[4] Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets [C]// Proc of the 28th Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2672–2680.

[5] Chen Shuyyu, Su Wanchao, Gao Lin, *et al.* Deepfacedrawing: deep generation of face images from sketches [J]. ACM Trans on Graphics, 2020, 39 (4): 72: 1–72: 16.

[6] Xiao Chufeng, Yyu Deng, Han Xiaoguang, *et al.* Sketchhairsalon: deep sketch-based hair image synthesis [J]. ACM Trans on Graphics, 2021, 40 (6): 216: 1–216: 16.

[7] Isola P, Zhu Junyan, Zhou Tinghui, *et al.* Image-to-image translation with conditional adversarial networks [C]// Proc of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 1125–1134.

[8] Yyu Qian, Liu Feng, Song Yizhe, *et al.* Sketch me that shoe [C]// Proc of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 799–807.

[9] Sangkloy P, Burnell N, Ham C, *et al.* The sketchy database: learning to retrieve badly drawn bunnies [J]. ACM Trans on Graphics, 2016, 35 (4): 119: 1–119: 12.

[10] Liu Ziwei, Luo Ping, Wang Xiaogang, *et al.* Deep learning face attributes in the wild [C]// Proc of the 2015 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2015: 3730–3738.

[11] Lu Yongyi, Wu Shangzhe, Tai Y, *et al.* Image generation from sketch constraint using contextual gan [C]// Proc of the 2018 European Conference on Computer Vision. Berlin, Springer Press, 2018: 205–220.

[12] Wah C, Branson S, Welinder P, *et al.* The Caltech-UCSD Birds-200-2011 dataset [EB/OL]. (2011) [2022-01-01]. citeseerx. ist. psu. edu/viewdoc/download; jsessionid=374669091E8C13903183C647B249A20B?doi=10. 1. 1. 372. 852&rep=rep1&type=pdf

[13] Krause J, Stark M, Deng Jia, *et al.* 3D object representations for fine grained categorization [C]// Proc of the 2013 IEEE International Conference on Computer Vision Workshops. Washington: IEEE Computer Society Press, 2013: 554–561.

[14] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks [C]// Proc of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2019: 4401–4410.

[15] Tan Zhentao, Chai Menglei, Chen Dongdong, *et al.* Michigan: multi-input-conditioned hair image generation for portrait editing [J]. ACM Trans on Graphics, 2020, 39 (4): 95: 1-95: 13.

[16] Chen Shuyyu, Liu Fenglin, Lai Yyukun, *et al.* Deepfaceediting: deep face generation and editing with disentangled geometry and appearance control [J]. ACM Trans on Graphics, 2021, 40 (4): 90: 1-90: 15.

[17] Wang Xiaogang, Tang Xiaoou. Face photo-sketch synthesis and recognition [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2009, 31 (11): 1955–1967.

[18] Xia Weihao, Yang Yyujiu, Xue Jinghao. Cali-sketch: stroke calibration and completion for high-quality face image generation from poorly-drawn sketches [EB/OL]. (2019-11-01) [2022-1-13]. https://doi. org/10. 48550/arXiv. 1911. 00426.

[19] Lee C, Liu Ziwei, Wu Lingyun, *et al.* Maskgan: towards diverse and interactive facial image manipulation [C]// Proc of the 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2020: 5548-5557.

[20] Li Yyuhang, Chen Xuejin, Yang Binxin, *et al.* Deepfacepencil: creating face images from freehand sketches [C]// Proc of the 28th International Conference on Multimedia. New York: ACM Press, 2020: 991–999.

[21] Karras T, Aila T, Laine S, *et al.* Progressive growing of gans for improved quality, stability, and variation [C/OL]/ Proc of the 6th International Conference on Learning Representations. (2018) [2022-01-01]. https://arxiv. org/abs/1710. 10196.

[22] Richardson E, Alaluf Y, Patashnik O, *et al.* Encoding in style: a stylegan encoder for image-to-image translation [C]// Proc of the 2021 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2021: 2287-2296.

[23] Liu Bingchen, Zhu Yizhe, Song Kunpeng, *et al.* Self-supervised sketch-to-image synthesis [C]// Proc of the 35nd AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2021: 2073-2081.

[24] Caesar H, Uijlings J, Ferrari V. Coco-stuff: thing and stuff classes in context [C]// Proc of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 1209–1218.

[25] Gao Chengying, Liu Qi, Xu Qi, *et al.* Sketchycoco: image generation from freehand scene sketches [C]// Proc of the 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2020: 5174–5183.

[26] Eitz M, Hays J, Alexa M. How do humans sketch objects? [J]. ACM Trans on Graphics, 2012, 31 (4): 44: 1–44: 10.

[27] Ha D, Eck D. A neural representation of sketch drawings [C/OL]/ Proc of the 6th International Conference on Learning Representations. (2017) [2022-01-01]. https://arxiv. org/abs/1704. 03477.

[28] Liu Runtao, Yyu Qian, Yu S. Unsupervised sketch-to-photo synthesis [C]// Proc of the 16th European Conference on Computer Vision. Berlin, Springer Press: 2020: 36-52.

[29] 宗雨佳. 两阶段草图至图像生成模型与应用实现 [D]. 大连: 大连理工大学, 2021. (Zong Yyujia. A two-stage method and application implementation for image generation from sketch [D]. Dalian: Dalian University of Technology. 2021.)

[30] Nilsback M, Zisserman A. Automated flower classification over a large number of classes [C]// Proc of the 6th Indian Conference on Computer Vision, Graphics&Image Processing. Washington: IEEE Computer Society Press, 2008: 722-729.

[31] 蔡雨婷, 陈昭炯, 叶东毅. 基于双层级联 GAN 的草图到真实感图像的异质转换 [J]. 模式识别与人工智能, 2018, 31 (10): 877-886. (Cai Yyuting, Chen Zhaojiong, Ye Dongyi. Bi-level cascading GAN-based heterogeneous conversion of sketch-to-realistic images [J]. Pattern Recognition and Artificial Intelligence, 2018, 31 (10): 877-886.)

[32] Qiu Haonan, Wang Chuan, Zhu Hang, *et al.* Two-phase hair image synthesis by self-enhancing generative model [J]. Computer Graphics

Forum, 2019, 38 (7): 403–412.

[33] Xie Saining, Tu Zhuowen. Holistically-nested edge detection [C]// Proc of the 2015 IEEE International Conference on Computer Vision. Washington: IEEE Computer Society Press, 2015: 1395–1403.

[34] Winnem H, Kyprianidis, J E, Olsen S. Xdog: an extended difference-of-gaussians compendium including advanced image stylization [J]. Computers & Graphics, 2012, 36 (6): 740 – 753.

[35] Kang H, Lee S, Chui C. Coherent line drawing [C]// Proc of the 5th International Symposium on Non-Photorealistic Animation and Rendering. New York: ACM Press, 2007: 43–50.

[36] Li Yijun, Chen Fang, Hertzmann A, *et al.* Im2pencil: controllable pencil illustration from photographs [C]// Proc of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2019: 1525–1534.

[37] Li Mengtian, Lin Zhe, Mech R, *et al.* Photosketching: inferring contour drawings from images [C]// Proc of the 2019 IEEE Winter Conference on Applications of Computer Vision. Piscataway, NJ: IEEE Press, 2019: 1403-1412.

[38] 刘波. 民族服饰草图自动着色方法研究 [D]. 云南: 云南师范大学, 2020. (Liu Bo. Automatic coloring method for national costume sketches [D]. Yunnan: Yunnan Normal University, 2020.)

[39] Photocopy. [2022-01-01]. Create filter gallery photocopy effect with single step in photoshop. https://www. youtube. com/watch?v=QNmniB_5Nz0.

[40] Sketch master. [2022-01-01]. http://www: ouyaoxiazai. corn/soft/txtx/108/8389. htm1.

[41] 王鹏程. 基于感知注意力和隐空间正则化的 GAN 在草图到真实图像的转换研究 [D]. 安徽: 安徽大学, 2020. (Wang Pengcheng. Research on gan translation from sketch to real image based on perceptual attention and latent space [D]. Anhui: Anhui University, 2020.)

[42] Ghosh A, Zhang R, Dokania P, *et al.* Interactive sketch&fill: multi-class sketch-to-image translation [C]// Proc of the 2019 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2019: 1171–1180.

[43] AutoTrace. [2022-01-01]. http://autotrace. sourceforge. net/.

[44] Simo-Serra E, Iizuka S, Sasaki K, *et al.* Learning to simplify: fully convolutional networks for rough sketch cleanup [J]. ACM Trans on Graphics, 2016, 35 (4): 121: 1–121: 11.

[45] Kyprianidis J E, Kang H. Image and video abstraction by coherence-enhancing filtering [J]. Computer Graphics Forum, 2011, 30 (2): 593-602.

[46] Olszewski K, Ceylan D, Xing Jun, *et al.* Intuitive, interactive beard and hair synthesis with generative models [C]// Proc of the 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2020: 7446–7456.

[47] Mirza M, Osindero S. Conditional generative adversarial nets [EB/OL]. (2014) [2022-01-01]. https://arxiv. org/abs/1411. 1784.

[48] Wang Tingchun, Liu Mingyyu, Zhu Junyan, *et al.* High-resolution image synthesis and semantic manipulation with conditional gans [C]// Proc of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 8798–8807.

[49] Mescheder L, Geiger A, Nowozin S. Which training methods for gans do actually converge? [C]// Proc of the 35th Annual International Conference on Machine Learning. New York: ACM Press, 2018: 3478-3487.

[50] Huang Xun, Liu Mingyu, Belongie S, *et al.* Multimodal unsupervised image-to-image translation [C]// Proc of the 15th European Conference on Computer Vision. Berlin, Springer Press, 2018: 172–189.

[51] Xiang Xiaoyyu, Liu Ding, Yang Xiao, *et al.* Adversarial open domain adaption for sketch-to-photo synthesis [C/OL]/ Proc of the 2022 IEEE Conference on Applications of Computer Vision. (2021) [2022-01-01]. https://arxiv. org/abs/2104. 05703.

[52] Karras T, Laine S, Aittala M. Analyzing and improving the image quality of stylegan [C]// Proc of the 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2020: 8107–8116.

[53] Zhu Junyan, Park T, Isola P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// Proc of the 2017 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2017: 2242–2251.

[54] Kim J, Kim M, Kang H, *et al.* U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation [C/OL]/ Proc of the 8th International Conference on Learning Representations. (2020-04-08) [2022-01-01]. https://arxiv. org/abs/1907. 10830.

[55] Huang Xun, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization [C]// Proc of the 2017 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2017: 1510–1519.

[56] Fu Jun, Liu Jing, Tian Haijie, *et al.* Dual attention network for scene segmentation [C]// Proc of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2019: 3146–3154.

[57] Kingma D P, Welling M. Auto-encoding variational bayes [C/OL]/ Proc of the 2nd International Conference on Learning Representations. (2013-12-20) [2022-01-01]. https://arxiv. org/abs/1312. 6114.

[58] He Kaiming, Fan Haoqi, Wu Yyuxin, *et al.* Momentum contrast for unsupervised visual representation learning [C]// Proc of the 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2020: 9729–9738.

[59] Liu Bingchen, Song Kunpeng, Zhu Yizhe, *et al.* Sketch-to-art: synthesizing stylized art images from sketches [C]// Proc of the 15th Asian Conference on Computer Vision. Berlin, Springer Press, 2020: 207-222.

[60] Wang Shengyyu, Bau D, Zhu Junyan. Sketch your own gan [C]// Proc of the 2021 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2021: 14030-14040.

[61] Zhang Pan, Zhang Bo, Chen Dong, *et al.* Cross-domain correspondence learning for exemplar-based image translation [C]// Proc of the 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2020: 5143–5153.

[62] Lee J, Kim E, Lee Y, *et al.* Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence [C]// Proc of the 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2020: 5800-5809.

[63] Lucic M, Kurach K, Michalski M, *et al.* Are gans created equal?a large-scale study [C]// Proc of the 2018 Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2018: 698–707.

[64] Heusel M, Ramsauer H, Unterthiner T, *et al.* Gans trained by a two time-scale update rule converge to a local Nash equilibrium [C]// Proc of the 2017 Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 6626-6637.

[65] Zhang R, Isola P, Efros A, *et al.* The unreasonable effectiveness of deep features as a perceptual metric [C]// Proc of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 586–595.

[66] Salimans T, Goodfellow I, Zaremba W, *et al.* Improved techniques for

training gans [C]// Proc of the 2016 Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2016: 2226–2234.

[67] Eitz M, Richter R, Boubekeur T, *et al.* Sketch-based shape retrieval [J]. ACM Trans on Graphics. 2012, 31 (4): 31: 1–31: 10.

[68] Li Yaoyi, Lu Hongtao. Natural image matting via guided contextual attention [C]// Proc of the 34th Conference on American Association for Artificial Intelligence. New York, AAAI Press, 2020: 11450-11457.

[69] Wang Zhou, Bovik A, Sheikh H, *et al.* Image quality assessment: from error visibility to structural similarity [J]. IEEE trans on image processing. 2004, 13 (4): 600–612.

[70] Xu Peng, Hospedales T, Yin Qiyue. Deep learning for free-hand sketch: a survey [J/OL]. IEEE Trans on Pattern Analysis and Machine Intelligence. (2020-06-01) [2022-01-01]. http://doi. org/10. 1109/TPAMI. 2022. 3148853.